INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Frequency and Spatial Domain-Based Approaches for Recognition of Indian Sign Language Gestures

B V Poornima[1]*, **S Srinath**[1]

1 Department of CSE, SJCE, JSS Science and Technology University, Mysuru, Karnataka, India

OPEN ACCESS

Received: 09-11-2023

Accepted: 23-12-2023

Published: 14-02-2024

Citation: Poornima BV, Srinath S (2024) Frequency and Spatial Domain-Based Approaches for Recognition of Indian Sign Language Gestures. Indian Journal of Science and Technology 17(7): 660-669. https://doi.org/10.17485/IJST/v17i7.2836

* **Corresponding author**.

poornimabv.85@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: The objective of this paper is to introduce and demonstrate an innovative approach for the recognition of Indian sign language gestures, with a focus on bridging communication gap between the deaf and hearing communities. The goal is to contribute to the development of effective tools and technologies that facilitate seamless communication between individuals using sign language and the people with no knowledge about sign language. **Methods:** The methodology consists of three key steps. First, data pre-processing involves resizing and contours extraction. Next, feature extraction employs Fourier descriptors for frequency domain analysis and gray-level-co-occurrence matrix for spatial domain analysis. Finally, various machine learning models including SVM, Random Forest, Logistic Regression, K-Nearest Neighbor and Naive Bayes are trained on a standard dataset. **Findings:** In our controlled experimental setup, we applied a diverse set of machine learning classifiers to evaluate the proposed approach for gesture recognition. Among the classifiers tested, K-Nearest Neighbors demonstrated the highest accuracy, achieving 99.82%. To validate the robustness of our approach, we employed k-fold cross-validation with 5 folds. **Novelty:** This study presents an innovative method for sign language recognition by employing a dual-domain fusion strategy that prominently emphasizes the frequency domain. Through the integration of Fourier descriptors, the research conducts a detailed frequency domain analysis to characterize the contour shapes of sign language gestures. The synergy with gray-level co-occurrence matrix texture features in the spatial domain analysis, contributes to the creation of a comprehensive feature vector. The proposed approach ensures a thorough exploration of gesture features, there by advancing the precision and efficacy of sign language recognition.

Keywords: Indian Sign Language (ISL); Sign Language Recognition (SLR); Frequency domain; Spatial domain; Fourier descriptors; Gray level cooccurrence matrix (GLCM); K Fold

1 Introduction

The hearing-impaired community in India relies on ISL as a visual means of communication, utilizing hand gestures, facial expressions, and the body motions. In the evolving landscape of technology, machine learning has emerged as an effective tool to bridge the communication gap between individuals with speech and hearing impairments and those without such impairments, facilitating the recognition and interpretation of gestures. Creating a machine learning model for gesture recognition involves pulling out important details from video or image data, like hand shape, position, finger placement, and color histograms. Numerous approaches in the literature address the challenge of SLR through various vision-based methods, utilizing diverse features of gestures essential for model training. Spatial information, encompassing the arrangement, position, and structure of objects in visual data, plays a pivotal role in recognizing complex visual patterns like sign language gestures[1]. In contrast, the frequency domain method represents the frequency content of sign language gestures, employing Fourier descriptors. This research seeks to enhance the effectiveness of ISL gesture recognition by integrating both spatial and frequency-based features using machine learning classifiers. While previous efforts predominantly focused on spatial features like local features, texture, color, and gradient features, these alone proved insufficient for SLR due to the diverse nature of sign language encompassing variations in hand shape, orientation, angles, skin color, and lighting conditions. In response to these challenges, this paper proposes a novel approach that combines Fourier descriptors with the GLCM to extract both global and local features, providing a more accurate recognition of signs. Despite the prevalence of reported SLR systems using deep learning models their limitations include the need for extensive data and high computational resources. While spatial domain features have demonstrated good performance, they fall short in extracting crucial dominant features present in ISL gestures. This research addresses these limitations by integrating both frequency and spatial domain approaches, focusing on extracting essential dominant features from images to enhance the accuracy of the classifiers. The decision to prioritize traditional machine learning methodologies over deep learning for the analysis of a dataset featuring simple images with a consistent black background is grounded in a thoughtful evaluation of the dataset's characteristics. The uniformity of the images suggests that the intricate feature extraction capabilities inherent in neural networks may be unnecessary for the task at hand. Instead, the simplicity of ML algorithms offers computational efficiency and interpretability, aligning with the observed straightforward relationships within the dataset. This strategic choice reflects a conscientious approach to selecting the most suitable methodology based on the specific properties and demands of the dataset, contributing to a more streamlined and interpretable solution, as detailed in this research paper. Beyond the intricacies of technology, the practical applications of SLR span various crucial aspects of life, significantly contributing to inclusivity and accessibility for the speech and hearing-impaired communities. In educational settings, the implementation of SLR fosters a more inclusive learning environment, empowering impaired students to engage more effectively with educational content. Similarly, in healthcare, the integration of this technology enhances communication between healthcare professionals and hearing-impaired patients, ensuring accurate and sensitive medical consultations. This has wide-reaching implications for the overall healthcare experience and outcomes. Moreover, in everyday social interactions, technology-driven SLR facilitates smoother communication for the hearing-impaired, allowing them to actively participate in a broader range of community engagements. Beyond these domains, SLR holds promise in employment opportunities, legal proceedings, public safety, entertainment and media accessibility, public transportation, and telecommunications. By addressing the technical challenges, this research not only advances technology but also has the potential to significantly improve the quality of life for the hearing-impaired population, making meaningful contributions across diverse practical applications in the real world. Figure 1 represents the overview of the proposed work architecture.

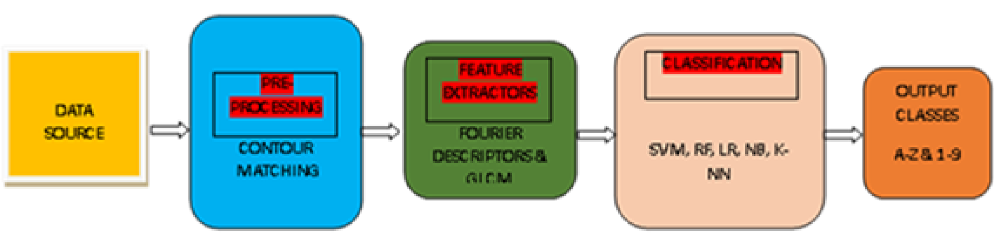

Fig 1. Overview of the proposed work

Dolly Indra et al.[2] introduced a method for BISINDO letter recognition centered on hand shape features, achieving remarkable accuracy exceeding 95%. Their approach involves shape segmentation, precise contour detection, and the extraction of features like contour count and chain code occurrence probability. However, a limitation emerges in the sensitivity of the employed chain code contour analysis and euclidean distance metric to variations in hand positioning, orientation, and scale. This may compromise accuracy in recognizing BISINDO letters with diverse hand shapes, highlighting the need for a more robust approach to accommodate such variations. Yang et al.[3] introduced an innovative multiscale Fourier descriptor approach for shape retrieval, emphasizing triangular features. Their method addresses a limitation of existing descriptors by introducing the "multiscale triangle feature," capturing detailed and global shape information. While achieving an impressive 84% accuracy in shape retrieval, a potential limitation lies in sensitivity to variations in scale and orientation. Careful parameter tuning is crucial for optimal performance, and guidelines for handling such variations could enhance the approach's practical applicability. While Deepali et al.'s[4] approach effectively addresses challenges posed by non-uniform backdrop datasets and achieves a commendable 95% accuracy, one potential limitation of their methodology could be its sensitivity to variations in skin tones. Since the approach relies on skin thresholding as a key step, it may be more susceptible to inaccuracies or reduced performance when applied to images with diverse skin tones. Skin tones can vary significantly across individuals and populations, and an approach heavily dependent on a specific threshold for skin detection might struggle to generalize well across diverse datasets. Therefore, the method's effectiveness could be influenced by the ability to adapt to a broad range of skin tones, and considerations for improving robustness in this aspect would be valuable for enhancing its applicability. Tavari et al.[5] introduced a method using histogram of oriented gradient (HOG) for recognizing Indian sign language hand gestures. The system identified 36 gestures representing alphabets A to Z and numbers 0 to 9. They employed an artificial neural network for classification and pre-processed images with a Gaussian low pass filter on a white background. The approach involved gradient computation, orientation binning, and histogram generation to extract HOG features. Gamal Tharwat et al.'s dataset[6] consisted of 9240 images. These images, which represent a range of ages and hand sizes for the gestures, were taken in ten different places. Various dimensions, angles, and intricate backdrops in the dataset caused difficulties. The study used statistical classifiers like C4.5, Nave Bayes, KNN, and a Multi-Layer Perceptron (MLP) neural network for gesture categorization and was able to achieve a hand identification accuracy of up to 98.64%. Shape-based descriptions were chosen, employing contour or region-based methods. The research encompassed several image enhancement techniques, including segmentation, morphological processing for noise reduction, and hand shape detection using three methods: Threshold, sobel filter for hand edge detection, and an estimation error classification experiment. Additionally, the impact of hidden neurons on MLP accuracy and the performance evaluation of classifiers were explored in various experiments. By taking a real-time photograph of the sign and translating it to its text counterpart, Nishi Intwala et al.'s system[7] employs the convolutional neural network technique to identify and classify the 26 letters of the Indian sign language into their corresponding alphabet letters. Here, the picture segmentation method GrabCut and the image classification algorithm MobileNet are both used. Results showed that run time image precision was 87.69% and testing picture precision was 96%. Wang[8] created features with peaks and valleys based on the slope difference distribution of the contour points. The resilience and accuracy for retrieving the peak and valley values are susceptible to being hampered by different types of noise. The positions of 20 joint points needed to be extracted from the depth map in accordance with the properties of the hand model were proposed as different sorts of features in[9], including joint rotation and fingertip distances. Their computational complexity is obviously great, and it is difficult to manage the extraction precision. Scale, translation, and rotation invariants are desirable for the feature descriptor in real-world applications[10]. As an illustration, the Hu moment invariants for hand gesture recognition were computed in after the contour of the hand region was retrieved with the Moore neighbour algorithm and the convex hull by the Graham scan approach. The computing load on this technique is high, though, and it is noise sensitive.

1.1 Observations made from the related work:

- Loss of detail in chain code contour analysis: Chain code simplifies shapes by representing contours as a sequence of directions. While efficient, it may fail to capture fine details and complex variations in shapes.
- Computational intensity of multiscale Fourier descriptors: Multiscale Fourier descriptor approaches involve complex mathematical operations, potentially resulting in longer processing times, especially with large or high-resolution datasets.
- Limited dataset size and diversity: Small datasets, like the one with 192 images, can lead to overfitting and may not represent real-world diversity, impacting the algorithm's performance.
- Limitation of HOG features: HOG features focus on local gradient information and may not capture higher-level spatial relationships or context, which can affect performance in scenarios where context is crucial.
- Specificity of the ISL system: Systems designed for specific sign languages, like the ISL system focusing only on 26 letters, lack versatility for broader sign languages or gestures.

- DTW's computational demands: Dynamic time warping (DTW) can be computationally intensive, which poses challenges for real-time applications and systems with strict latency requirements.
- Scalability of joint points for gesture recognition: Using joint points for feature extraction may not scale effectively for recognizing a large number of hand gestures or complex vocabularies. These drawbacks highlight common challenges in various gesture recognition methods, emphasizing the need for robustness, scalability, and efficiency in such systems.

1.2 Research gap

- Existing methods predominantly rely on spatial domain representations, neglecting the potential advantages offered by frequency-based analyses, such as Fourier or wavelet transforms. Progress in the field of sign language recognition has to be made by determining how to use these frequency features to improve the systems' ability to recognize signs.

1.3 Contribution of the proposed system

- In this study, we investigate both the frequency and spatial domains to present a unique method for ISL gesture recognition. Combining Fourier descriptors (FD) for contour shapes with gray-level co-occurrence matrix (GLCM) texture characteristics allows the approach to capture fine gesture details.
- The integration of both frequency and spatial domain-based feature extraction approaches is motivated by their complementary nature in capturing distinct characteristics of static Indian Sign Language (ISL) gestures.
- The frequency domain excels in revealing fine details related to the spatial configurations and shapes of the static gestures, providing insights into the inherent patterns and textures. Simultaneously, the spatial domain emphasizes the specific static positioning and structural attributes of the gestures.
- Using this dual-domain approach yields a robust feature vector that improves the accuracy of recognition.

The fusion approach incorporating FD and GLCM in the recognition of ISL alphabets and digits combines the extraction of frequency and spatial features. FD plays a key role in capturing frequency domain information by decomposing the spatial shape of the hand into sine and cosine components, providing insights into the static form and structural characteristics of the gestures. On the other hand, GLCM focuses on spatial relationships between pixel intensities, calculating the frequency of pixel pairs at various distances and angles. This spatial feature extraction method is essential for capturing textural and structural information within static sign language gestures. By integrating both types of features, the fusion approach ensures a comprehensive representation, leveraging the complementary nature of frequency and spatial details for accurate recognition of static sign language expressions.

2 Methodology

2.1 Dataset details

The research is based on a Kaggle[(11)] benchmark dataset containing images ISL alphabets (A-Z) and numbers (1-9) in a plain black background. Notably, each distinct sign within the dataset has a set of 1000 images. ISL digits are represented using single hand and alphabets are represented using double hands. Total volume of the dataset is 35000. The initial resolution of the images was 128x128 pixels, and as part of the pre-processing stage, the images were subsequently resized to a lower dimension.

2.2 Pre-processing

Data pre-processing is one of the crucial steps to prepare the data to make it suitable for faster execution. In our research, we employed contour detection technique as a pivotal step in extracting regions of interest (ROI). This involved pre-processing the images by converting them to grayscale and applying relevant filters for feature enhancement. Subsequently, we utilized contour matching[(12)] algorithm to identify and delineate object boundaries within the images. The identified contours enable the extraction of specific ROI. This approach, applied to images with a simple black background, demonstrated its efficacy in isolating meaningful features and contributing to the overall success of our image analysis methodology. The images are captured in different angle and orientation. Resizing, rotating, and scaling of images have been implemented to evaluate the robustness of the proposed model. Figure 2 shows the samples of the dataset.

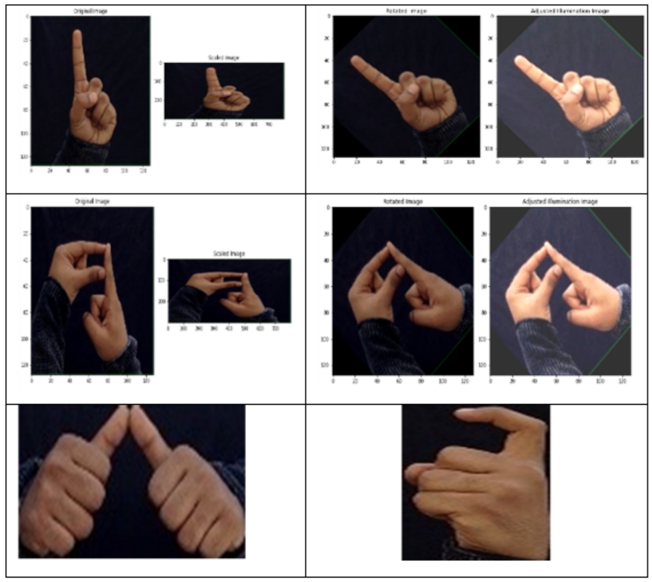

Fig 2. Dataset samples

2.3 Feature extraction

2.3.1 ***Fourier descriptors***

Derived from the Fourier transform, the Fourier descriptors encode the spatial information of an object's contour into a set of complex coefficients, revealing the dominant frequencies and their relative strengths in the object's boundary. Notably, Fourier descriptors possess inherent rotation and scale invariance, making them robust in recognizing shapes irrespective of orientation or size changes. Figure 3 represents the steps in extracting feature vectors using Fourier descriptors.

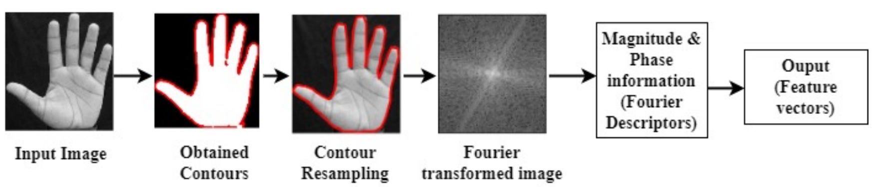

Fig 3. Feature extraction using Fourier descriptors

2.3.2 ***Details of the experiment***

- Contour Resampling

Contours of an image is resampled [13] to a specified number of points. It is done by removing any redundant dimensions from the contour. Then the length of the new contour obtained is calculated. The x and y coordinates of the contour points are extracted and interpolated linearly. The resulting resampled x and y coordinates are stacked together to form the resampled contour.

- Extract Fourier descriptors

Fourier descriptors are extracted from a given resampled contour which gives a consistent number of points. The fast Fourier transform (FFT) [14] is applied to the resampled contour to obtain complex Fourier coefficients. From the complex coefficients, magnitude and phase spectrum are calculated.

- Feature vector creation: The magnitude and phase of the complex Fourier coefficients are concatenated into a single feature vector. The magnitude captures the relative importance of different frequencies, while the phase represents the spatial arrangement of the frequencies. Inverse Fourier transformation is not required in this context because the classification is based on the Fourier descriptors directly.
- **An overview of the process involved in calculating Fourier descriptors**
  - Collect shape signature: Obtain the (x, y) coordinates of the object's external boundary.
  - Convert to complex coordinates: Convert the (x, y) coordinates into complex coordinates using Equation: S(k)=X(k)+jY(k).
  - Apply Fourier transform: The Fourier transform is applied to the complex coordinates of the shape signature.

• Analyze Fourier coefficients: The complex coefficients [15] obtained from the Fourier transform are the Fourier descriptors. These coefficients capture the frequency components of the shape.

*2.3.3 **GLCM***

GLCM captures the spatial relationships [16] between pixel values (gray levels) in an image, providing statistical information about how frequently certain pairs of pixel values occur together within a specific neighborhood. This matrix can then be analyzed to extract various texture-related features that describe the patterns and relationships present in the image.

- **Steps to perform texture analysis**
  - Apply gabor filter (Choose Gabor filter parameters: filter size, orientation, wavelength, standard deviation, phase offset, and aspect ratio). Figure 4 shows the filtered image using Gabor filer [17].

```
# Gabor filter parameters
    ksize = 3 # Filter size
    sigma = 4.0 # Standard deviation
    theta = np.pi/4 # Orientation in radians
    lambda_ = 10.0 # Wavelength
    psi = 0 # Phase offset
    gamma = 0.5 # Aspect ratio
```

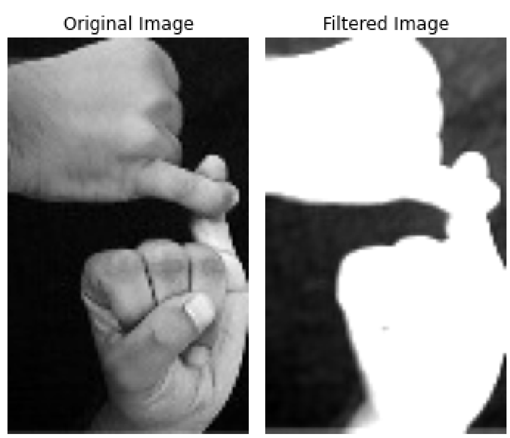

Fig 4. Filtered image

- Calculate GLCM

Define the properties of the GLCM: distance and angles between pixel pairs, and the number of gray levels. For each pixel in the image, calculate the co-occurrence matrix by counting the occurrences of pixel pairs with the defined properties.

- Extract GLCM features

Calculate texture features from the normalized GLCM matrix. Common features include contrast, energy, correlation, and homogeneity. These features capture different aspects of the texture patterns present in an image.

Contrast: [[1781.8333872 4489.81895425 1704.13719807 296.03153595]]

Energy: [[0.57346333 0.54584707 0.57555821 0.5964472]]

Correlation: [[0.90661944 0.76423282 0.91074641 0.9844527]]

Homogeneity: [[0.66396515 0.61445073 0.6646186 0.72278485]].

• Flattening the features

Concatenate the extracted GLCM features (contrast, energy, correlation, homogeneity) with the Fourier descriptors. This creates a combined feature vector that includes both texture-related features and shape-related features.

2.4 Classification

Based on the notion of probability or distance measurements, classification divides input vectors into distinct classes. There are 35 different classes represented in the proposed research. Feature vectors collected from 28000 of the 35000 input images are sent to the classifier as training samples. As test samples, the remaining feature vectors from the 7000 images are employed. Consequently, the train to test ratio is 80:20. The training process, was conducted on jupyter notebook utilizing a high-end GPU. The total duration required for training was approximately 20 hours. The classifiers are able to forecast the class labels of testing vectors based on the created model and testing samples. Experiment was also evaluated using various classifiers like support vector machine (SVM), random forest (RF), Naïve Bayes (NB), logistic regression (LR). K-Nearest Neighbor (K-NN) has given better accuracy compared to other classifiers.

3 Results and Discussion

Experiment is carried out using the concatenation of frequency and spatial domain feature extraction technique in conjunction with various classifiers, as discussed in the previous section. Figure 5 shows the average recognition rate obtained by the various classifiers used for the experimentation. Table 1 shows the character wise results obtained by K-NN classifier. The performance evaluation metrics used are precision, recall and F1 score, and accuracy and the results are represented in Table 2. Cross validation is done using K-fold method to validate the model's performance. The results are represented in Table 3. Figure 6 represents the results in the form of confusion matrix.

Table 1. Character wise recognition rate obtained by K-NN classifier

ISL SIGN	Accuracy	ISL SIGN	Accuracy	ISL SIGN	Accuracy	ISL SIGN	Accuracy
1	0.98	A	0.99	J	0.97	S	0.99
2	0.96	B	0.98	K	0.96	T	0.96
3	0.97	C	1.00	L	0.94	U	0.96
4	0.99	D	0.95	M	0.98	V	0.99
5	1.00	E	0.97	N	0.98	W	0.96
6	0.99	F	0.98	O	0.93	X	0.98
7	1.00	G	0.97	P	0.94	Y	0.99
8	0.99	H	0.98	Q	0.95	Z	0.96
9	0.98	I	1.00	R	0.97	Average %= 97.4	

The recognition rates for the set of 35 signs exhibit variability, with certain gestures achieving a 100% recognition rate while others, specifically 'D,' 'L,' and 'P,' demonstrate relatively lower prediction rates. The observed disparities can be attributed to the inherent complexity of certain signs, the variability in gesture execution, the efficacy of the feature extraction process, and the sensitivity of the employed classifier. The challenges arising from intricate movements, gesture variations, and the classifier's discernment capabilities contribute to the divergent recognition outcomes.

Table 2. Performance evaluation metrics table for KNN Classifier

ISL Sign	Precision	Recall	F1 Score
Digits	0.97	0.98	0.98
Alphabets	0.98	0.98	0.99

Table 3. Quantitative assessment using cross validation (5 folds)

K-Fold method		
S NO.	ISL Signs	KNN
1	Digits	99.60
2	Alphabets	99.85

The implementation of 5-fold cross-validation in this study has been instrumental in enhancing the overall performance of the model. This technique involves dividing the dataset into five subsets, or "folds," and iteratively training the model on four folds while validating on the remaining fold. This process is repeated five times, each time using a different fold for validation. By averaging the results across these iterations, the model is exposed to a more comprehensive range of data during both training and validation phases. This mitigates the impact of potential biases introduced by a specific training-validation split, leading to a more robust evaluation of the model's generalization capabilities. Consequently, the adoption of 5-fold cross-validation serves to improve the reliability and accuracy of the performance assessment, providing a more representative indication of the model's effectiveness across various data scenarios. The superior accuracy attained by K-NN model is a result of meticulous parameter tuning, a process aimed at optimizing key parameters that profoundly influence the model's performance. We performed a systematic search for the optimal number of neighbors ('K') using techniques like grid search and cross-validation. Notably, after careful evaluation, we determined that setting 'k' to 5 yielded the highest accuracy on the validation set. In considering the distance metric, we explored both Euclidean and Manhattan distances, given the balanced nature of our dataset with an equal number of images in all classes. Cross-validation results revealed that the Manhattan distance metric was better suited to capture underlying patterns in our feature space, leading to improved model performance. Furthermore, even with a balanced dataset, we delved into different weighting schemes for neighbors, such as uniform weighting and distance-weighted voting. Notably, our experiments highlighted the effectiveness of distance-weighted voting in enhancing accuracy across all classes.

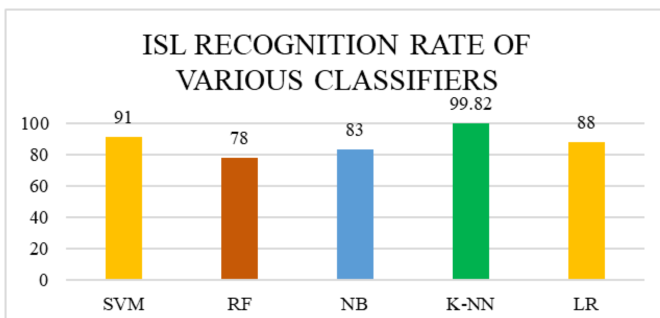

Fig 5. Recognition rate obtained by various classifiers

The selected frequency and spatial domain features are inherently computationally efficient. Their nature allows for straightforward calculations, mitigating concerns of excessive computational overhead. Despite the integration of these features, the computational demands remain reasonable. Preliminary empirical observations and experiments further support the efficiency of the integrated approach. Even without explicit optimizations, our method demonstrates reasonable computational performance. This observation aligns with the goal of maintaining practical feasibility for real-world applications. While the current implementation may not extensively optimize for computational complexity, it lays the groundwork for subsequent research phases to explore targeted optimizations without compromising the core integrated approach.

3.1 Comparative analysis

Table 4 shows that while existing systems have achieved commendable accuracy rates ranging from 98% to 99.71%, the proposed SLR method, incorporating Fourier descriptors, GLCM, and a K-NN classifier with 5-fold cross-validation, surpasses these achievements with an accuracy of 99.82%. This improvement can be attributed to the synergistic combination of Fourier descriptors and GLCM, capturing both frequency and spatial information, enhancing the discriminative power of the feature set. The inclusion of the K-NN classifier further contributes to the robustness of the model, allowing it to adapt well to diverse sign gestures. Moreover, the utilization of 5-fold cross-validation ensures a more comprehensive evaluation, minimizing the impact of dataset biases and enhancing the generalization capability of the model. The proposed method stands out not only

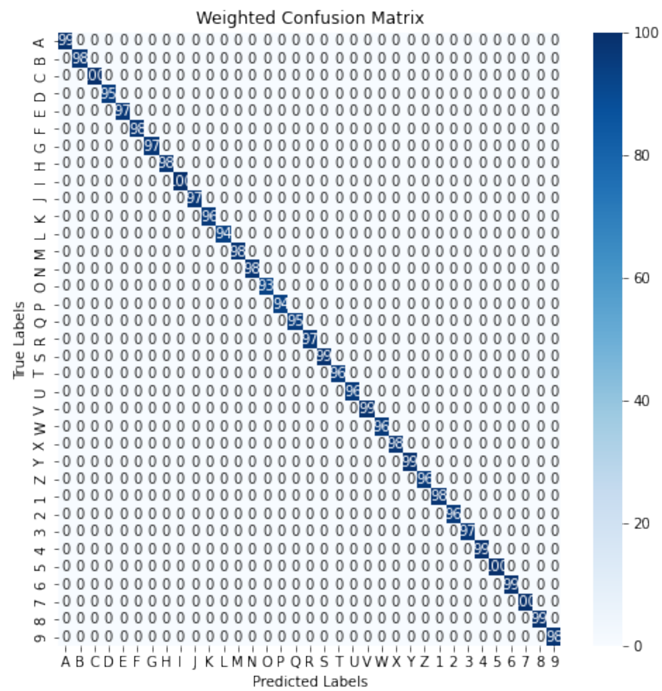

Fig 6. Confusion matrix

for its superior accuracy but also for its resilience to variations in sign execution, making it a promising advancement in the field of SLR.

Table 4. Existing systems Vs. Proposed method

S NO.	Year	Methodology	Accuracy
1	2020 [18]	Surf key points, SVM classifier	99%
2	2022 [19]	SIFT, SVM classifier	99.71%
3	2021 [20]	CNN	98%
4	2022 [21]	ORB key points, SVM classifier	99%
5	**Proposed**	**Fourier descriptors, GLCM, KNN classifier**	**99.82% (K-fold method)**

4 Conclusion

In conclusion, this research paper introduces a novel hybrid methodology for enhancing the recognition of ISL gestures, with a focus on facilitating improved communication between the speech and hearing communities. The proposed approach integrates Fourier descriptors (FD) in the frequency domain and gray-level co-occurrence matrix (GLCM) texture features in the spatial domain to comprehensively capture intricate shape and texture details, significantly elevating the accuracy of ISL gesture recognition. Experimental validation was conducted on a meticulously curated dataset, comprising gestures recorded against uniform black backgrounds, and explicitly addressing alphabets and digits within a controlled environment. The dataset was further enriched by including transformed images, such as scaled and rotated variations, to assess the robustness and generalizability of the proposed methodology demonstrates the effectiveness of our methodology. While the current research successfully achieves its predefined objectives, offering valuable insights and showcasing the superiority of the K-Nearest Neighbors (KNN) classifier, it is essential to recognize the constrained scope of our dataset. The confined focus on black backgrounds, along with a limited gestures like alphabets and digits, presents a limitation in capturing the diversity inherent in naturalistic settings. Future investigations should extend beyond this controlled environment, encompassing diverse backgrounds and a more comprehensive range of ISL gestures to enhance the generalizability and applicability of the proposed methodology in real-world scenarios. This research serves as a foundational step towards advancing assistive technologies for meaningful interactions between ISL users and those relying on spoken language.

References

1) Vazquez-Enriquez M, Alba-Castro JL, Docio-Fernandez L, Rodriguez-Banga E. Isolated Sign Language Recognition with Multi-Scale Spatial-Temporal Graph Convolutional Networks. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE. 2021;p. 3457–3466. Available from: https://doi.org/10.1109/CVPRW53098.2021.00385.
2) Indra D, Purnawansyah, Madenda SP, Wibowo EP. Indonesian Sign Language Recognition Based on Shape of Hand Gesture. *Procedia Computer Science*. 2019;161:74–81. Available from: https://doi.org/10.1016/j.procs.2019.11.101.
3) Yang C, Yu Q. Multiscale Fourier descriptor based on triangular features for shape retrieval. *Signal Processing: Image Communication*. 2019;71:110–119. Available from: https://doi.org/10.1016/j.image.2018.11.004.
4) Mali D, Limkar N, Mali S. Indian Sign Language Recognition using SVM Classifier. In: International Conference on Communication and Information Processing (ICCIP-2019). 2019;p. 1–9. Available from: https://dx.doi.org/10.2139/ssrn.3421567.
5) Tavari NV, Deorankar AV. Indian Sign Language Recognition based on Histograms of Oriented Gradient. *International Journal of Computer Science and Information Technologies*. 2014;5(3):3657–3660. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b188e93f8aa17a6c415bb2e558ee973538510f25#:~:text=For%20the%20recognition%20of%20ISL,the%20purpose%20of%20object%20detection.
6) Tharwat G, Ahmed AM, Bouallegue B. Arabic Sign Language Recognition System for Alphabets Using Machine Learning Techniques. *Journal of Electrical and Computer Engineering*. 2021;2021:1–17. Available from: https://doi.org/10.1155/2021/2995851.
7) Intwala N, Banerjee A, Meenakshi, Gala N. Indian Sign Language converter using Convolutional Neural Networks. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE. 2020;p. 1–5. Available from: https://doi.org/10.1109/I2CT45611.2019.9033667.
8) Wang ZZ. Gesture recognition by model matching of slope difference distribution features. *Measurement*. 2021;181:109590. Available from: https://doi.org/10.1016/j.measurement.2021.109590.
9) Miao Y, Li J, Liu J, Chen J, Sun S. Hand Gesture Recognition Based on Joint Rotation Feature and Fingertip Distance Feature. *Chinese Journal of Computers*. 2020;1:78–92. Available from: https://caod.oriprobe.com/articles/58986786/Hand_Gesture_Recognition_Based_on_Joint_Rotation_F.htm.
10) Dhiman C, Vishwakarma DK. A Robust Framework for Abnormal Human Action Recognition Using R -Transform and Zernike Moments in Depth Videos. *IEEE Sensors Journal*. 2019;19(13):5195–5203. Available from: https://doi.org/10.1109/JSEN.2019.2903645.
11) Karthikeyu, Shravanikatte. Indian-sign-language-recognition. . Available from: https://github.com/Karthikeyu/Indian-sign-language-recognition.
12) Subudhi P, Mukhopadhyay S. A statistical active contour model for interactive clutter image segmentation using graph cut optimization. *Signal Processing*. 2021;184:108056. Available from: https://doi.org/10.1016/j.sigpro.2021.108056.
13) Liu J, Yuan J, Cui J, Liu Y, Liu X. Contour Resampling-Based Garlic Clove Bud Orientation Recognition for High-Speed Precision Seeding. *Agriculture*. 2022;12(9):1–26. Available from: https://doi.org/10.3390/agriculture12091334.
14) Khan US, Zhang X, Su Y. Active Contour Model Using Fast Fourier Transformation for Salient Object Detection. *Electronics*. 2021;10(2):1–20. Available from: https://doi.org/10.3390/electronics10020192.
15) Kim T, Jung YH. Correlation of Roundness and Sphericity with Complex Fourier Descriptors of 2-Dimensional Projected Images of Soil Particles. *KSCE Journal of Civil Engineering*. 2022;26(3):1089–1094. Available from: https://doi.org/10.1007/s12205-021-5922-5.
16) Singh J, Singh H, Goyal V. A Comprehensive Study on Feature Extraction and Classification Techniques for Sign Language Recognition. In: 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE. 2022;p. 959–964. Available from: https://doi.org/10.1109/ICECA52323.2021.9676024.
17) Erwin, Karo NNB, Sari AY, Aziza N, Putra HK. The Enhancement of Fingerprint Images using Gabor Filter. In: International Conference on Information System, Computer Science and Engineering;vol. 1196 of Journal of Physics: Conference Series. IOP Publishing. 2019;p. 012045–012045. Available from: https://doi.org/10.1088/1742-6596/1196/1/012045.
18) Shravani K, Lakshmi AS, Geethika MS, Kulkarni SB. Indian Sign Language Character Recognition. *IOSR Journal of Computer Engineering (IOSR-JCE)*. 2020;22(3):14–19. Available from: https://www.iosrjournals.org/iosr-jce/papers/Vol22-issue3/Series-1/B2203011419.pdf.
19) Munnaluri V, Pandey V, Singh P. Machine Learning based Approach for Indian Sign Language Recognition. In: 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE. 2022;p. 1128–1132. Available from: https://doi.org/10.1109/ICCES54183.2022.9835908.
20) Indian Sign Language Recognition Using Canny Edge Detection. *International Journal of Advanced Trends in Computer Science and Engineering*. 2021;10(3):1576–1583. Available from: https://doi.org/10.30534/ijatcse/2021/131032021.
21) Savant R, Nasriwala J, Bhatt P. Static Gesture Recognition for Indian Sign Language Alphabets and Numbers using SVM with ORB Keypoints and Image Pixel As Feature. *International Journal of Engineering Research & Technology (IJERT)*. 2022;11(1):214–219. Available from: https://www.ijert.org/research/static-gesture-recognition-for-indian-sign-language-alphabets-and-numbers-using-svm-with-orb-keypoints-and-image-pixel-as-feature-IJERTV11IS010090.pdf.